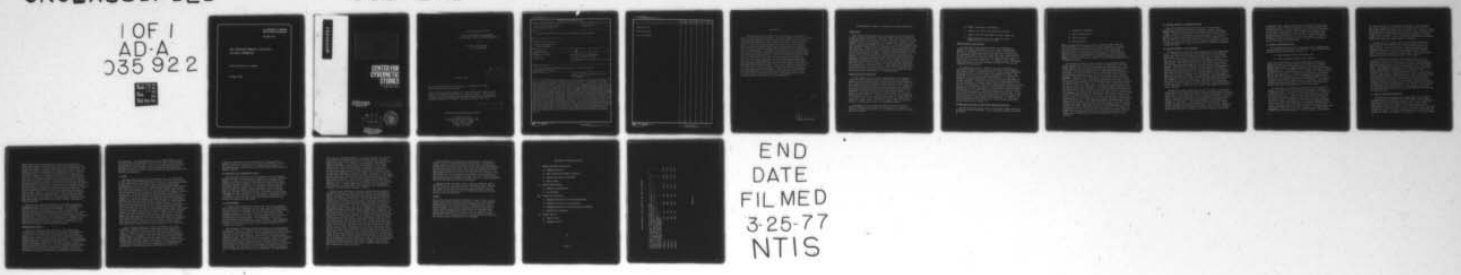
AD-A035 922 TEXAS UNIV AT AUSTIN CENTER FOR CYBERNETIC STUDIES F/G 9/2
DATA PROCESSING PROBLEMS IN COLLECTING JOB SURVEY INFORMATION.(U)
OCT 76 J J WEISMULLER, J B KAUFFMAN N00014-75-C-0616

UNCLASSIFIED CCS-278 NL

1 OF 1
AD-A
035 922



END
DATE
FILMED
3-25-77
NTIS

U.S. DEPARTMENT OF COMMERCE
National Technical Information Service

AD-A035 922

DATA PROCESSING PROBLEMS IN COLLECTING JOB SURVEY INFORMATION

TEXAS UNIVERSITY AT AUSTIN

OCTOBER 1976

ADA035922

CENTER FOR CYBERNETIC STUDIES

The University of Texas
Austin, Texas 78712

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U. S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

DDC
RECEIVED
FEB 23 1977

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U. S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

Research Report CCS 278

DATA PROCESSING PROBLEMS IN COLLECTING JOB SURVEY INFORMATION

by

Johnny J. Weismuller*
J. Britt Kauffman**

DDC
RECEIVED
FEB 23 1977
A

October 1976

*Air Force Human Resources Laboratory, Lackland AFB, Texas.
**The University of Texas at Austin.

This research was partly supported by Project NR047-021, ONR Contract N00014-75-C-0616 with the Center for Cybernetic Studies, The University of Texas, and was presented at the 18th Annual Conference of the Military Testing Association, Gulf Shores, Alabama, October 18-22, 1976.

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

CENTER FOR CYBERNETIC STUDIES

A. Charnes, Director
Business-Economics Building, 203E
The University of Texas
Austin, Texas 78712
(512) 471-1821

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R & D

Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified

1. ORIGINATING ACTIVITY (*Corporate author*) Center for Cybernetic Studies The University of Texas	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

"Data Processing Problems in Collecting Job Survey Information"

4. DESCRIPTIVE NOTES (*Type of report and inclusive dates*)

5. AUTHOR(S) (*First name, middle initial, last name*)

Johnny J. Weissmuller
J. Britt Kauffman

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
October 1976	19	0

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
N00014-75-C-0616 b. PROJECT NO NR047-021 c. d.	Center for Cybernetic Studies Research Report CCS 278 9b. OTHER REPORT NO(S) (*Any other numbers that may be assigned this report*)

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Office of Naval Research (Code 434) Washington, D. C.

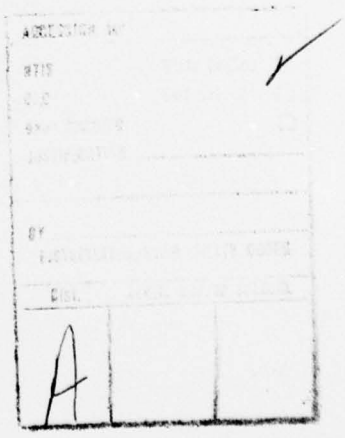
13. ABSTRACT

Many data processing problems may be avoided by the improved design of data collection instruments. Improved design, however, necessitates a thorough understanding of both the input requirements and the operating characteristics of the processing systems to be used. Normally it is not feasible for the person preparing the instrument to be totally familiar with the computer programming packages, and it is therefore essential that the instrument designer locate and communicate with the people who do have this knowledge. Depending on the diversity of processing required, this may entail identifying and contacting several distinct groups of people in the data processing department. When this communication fails to take place, design decisions are made on criteria which are thought to be arbitrary, but which actually have significant processing implications. Moreover, the resulting problems are not even anticipated by the data processing department who becomes aware of the problems only after the operational processing has begun. By this time, problems which would have been easy to correct or prepare for have become major obstacles in extracting the desired information. Major obstacles, though usually not insurmountable, do exact a high price not only in computer time, but also in stress-filled man-hours spent by the programmer and those frustrating delays experienced by the waiting analyst.

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Opscan Form						
Data Collection						
Data Processing						

ABSTRACT

Many data processing problems may be avoided by the improved design of data collection instruments. Improved design, however, necessitates a thorough understanding of both the input requirements and the operating characteristics of the processing systems to be used. Normally it is not feasible for the person preparing the instrument to be totally familiar with the computer programming packages, and it is therefore essential that the instrument designer locate and communicate with the people who do have this knowledge. Depending on the diversity of processing required, this may entail identifying and contacting several distinct groups of people in the data processing department. When this communication fails to take place, design decisions are made on criteria which are thought to be arbitrary, but which actually have significant processing implications. Moreover, the resulting problems are not even anticipated by the data processing department who becomes aware of the problems only after the operational processing has begun. By this time, problems which would have been easy to correct or prepare for have become major obstacles in extracting the desired information. Major obstacles, though usually not insurmountable, do exact a high price not only in computer time, but also in stress-filled man-hours spent by the programmer and those frustrating delays experienced by the waiting analyst.



DATA PROCESSING PROBLEMS IN COLLECTING JOB SURVEY INFORMATION

INTRODUCTION

Many data processing problems may be avoided by the improved design of data collection instruments. Improved design, however, necessitates a thorough understanding of both the input requirements and the operating characteristics of the processing systems to be used. Normally it is not feasible for the person preparing the instrument to be totally familiar with the computer programming packages, and it is therefore essential that the instrument designer locate and communicate with the people who do have this knowledge. Depending on the diversity of processing required, this may entail identifying and contacting several distinct groups of people in the data processing department. When this communication fails to take place, design decisions are made on criteria which are thought to be arbitrary, but which actually have significant processing implications. Moreover, the resulting problems are not even anticipated by the data processing department who becomes aware of the problems only after the operational processing has begun. By this time, problems which would have been easy to correct or prepare for have become major obstacles in extracting the desired information. Major obstacles, though usually not insurmountable, do exact a high price not only in computer time, but also in stress-filled man-hours spent by the programmer and those frustrating delays experienced by the waiting analyst.

GENERAL DESIGN CONSIDERATIONS

On the highest level, the prime considerations in the design of data collection instruments should be both economy and reliability. Normally in data processing clear trade-offs exist between these goals, but through technological advances coupled with good design, both of these objectives can be maximized without sacrifice. For example, in the area of occupational analysis many organizations are currently in the process of converting their initial data collection/entry from keypunched inventories to optically scanned response sheets. It is hoped that this change will reduce the total overall cost while not impairing the reliability. To insure that this hope can be realized, it is imperative that the criteria for the design of opscan sheets be considered at this time. A good design is only a good design with respect to a given context, and in this paper we deal only with the design of opscan sheets for use in occupational analysis.

Although we are restricting our context in order to provide specific examples, the principles of the analysis are general. To collect the information required for an adequate design, the following steps are required:

1. Define the context of the design,

2. Identify the systems and sub-systems involved,

3. Analyze the input requirements of each system, and

4. Seek information on both the ideal and actual operating characteristics of each.

SPECIFIC DESIGN CONSIDERATIONS

At this level in defining our context it is appropriate to identify the unique characteristics of job survey information as opposed to a knowledge test or an opinion survey. Historically, job survey instruments have evolved into task check lists in which job incumbents indicate only those tasks which they actually perform. In order to identify similarities across jobs, these check lists, or task inventories, grew in scope until an incumbent would respond to perhaps one out of every eight tasks in a typical survey. This methodology decreased the cost and increased the speed of collecting and processing job information.

The significant point is, however, that the job incumbent is selecting a relatively small number of items from the full inventory. This characteristic, low response density, creates unique problems and opportunities for both the processing and collection of data. For processing this type of data, computer programs can be written which capitalize on this sparse response characteristic. In fact, the package of computer programs with which we will deal is designed based on this assumption. We will be referring to the Comprehensive Occupational Data Analysis Programs (CODAP) system as developed by the U.S. Air Force Human Resources Laboratory and implemented by several government agencies and universities. Because the CODAP design is based on this assumption, it is very efficient for processing job survey information, but conversely it is rather cumbersome and costly to process a survey in which each respondent is required to answer every question. In the area of data collection this low response density can be either a time-saving advantage or a frustrating disadvantage, and the design of the data collection instrument itself makes the difference. This point will be elaborated in the discussion of various implementations and the Case History which follows it.

SYSTEMS AND SUB-SYSTEMS IN JOB SURVEY INFORMATION HANDLING

In the most general sense, there are four major systems involved in handling job survey information. These systems may roughly be defined as follows:

1. Instrument Construction

2. Survey Administration

3. Data Processing

4. Survey Analysis

Figure 1 demonstrates how each of these systems contains people who are responsible for the assumptions and daily operations. Key people in each of these systems can usually provide valuable information to aid in the design of the survey instruments. As data processing specialists, we, the authors, have been approached by various people within the data processing department regarding job survey instruments. Because of these initial contacts we decided to actively solicit more comments to be included here as a sample of data processing suggestions for increased efficiency.

In soliciting information about specific data processing problems, a survey designer may make an honest attempt at communication, but find himself talking with a totally disinterested (or confused) party. This is the result of the fact that non-data processing personnel usually fail to realize that there are distinct areas of specialization within data processing. At the very minimum there are four specialities involved in the processing occupational data. The first of these areas is that of designing the programs and procedures for use with the optical scanning equipment. Because this is a relatively new frontier, the person in this area will usually be a very sophisticated programmer/analyst who is used to accepting new challenges. The second area is that of actually operating the optical scanner on a daily basis. In this area, one finds people who are apt to know all the things which can retard processing; things totally unsuspected by the above programmer or the optical scan machine serviceman. The third specialization encompasses the programmer/analyst responsible for the design of the computer analysis programs. In our context, he would be the programmer who maintains and upgrades the CODAP system. The fourth area is that of operational studies. The people in this area routinely accept tapes or punch cards produced by the optical scan department, correct the detectable errors, and provide the survey analyst with the requested reports using the CODAP system. These CODAP applications programmers are usually familiar with conversion and utility programs outside of the CODAP system because the optical scanner tapes require pre-processing prior to entering the data into CODAP. Depending on the size of the organization, a single person may act as one or more of these four specialists. The larger the organization, the more likely that each area will have distinct people or groups. We contacted this range of data processing personnel to solicit observations or suggestions and the following sections contain their responses.

A. OPTICAL SCANNING: PROGRAMMER/ANALYSTS

From the programmer/analysts who design the optical scanner programs and conversion procedures we received a request to avoid mixing alphabetic and numeric characters together in the same response columns. If that is not possible, at the very least, point out this condition before the translation programs are written. Secondly, avoid splitting a single character response into multiple columns or rows. If, for example, a single digit response of 1 to 9 is solicited on three rows (1, 2, 3), (4, 5, 6), and (7, 8, 9) special programming is required to make the conversion program understand that the three rows represent a single response.

B. OPTICAL SCANNING: DAILY OPERATORS

The operators of the optical scan equipment had several suggestions. First, if specially designed and over-printed optical scan sheets are being used, test batches of these forms should be run on the scanner before these forms are mass produced and sent out to the field. These test runs are invaluable for detecting such problems as missing or improper page codes, improper trimming of the booklets, and unnecessary "dark marks". An excessive "dark mark" in the center of a scan sheet can cause some scanning equipment to halt and require operator intervention. Such "dark mark" conditions have been caused by organizational emblems printed on the cover sheets of survey booklets. The service personnel who do maintenance on the scanner point out that such massive dark areas overload the sensors and cause the machine to halt, but they seem to be at a loss to explain why a fine-lined "X" across an entire page would have the same effect. Well-intentioned survey designers have been known to have such "X" marks printed on pages along with instructions to incumbents to not mark on that page. Thus far we have been lucky that the reverse side of these offending pages has not contained data because the scanner attempts to read both sides of a sheet simultaneously, and the operators, to avoid the halts, remove these sheets prior to the actual scanning.

Though the following observations are not the result of the survey design itself, the optical scan machine operators believe that additional instructions on the booklets may help to reduce these problems. The second item which can delay processing is that of careless handling of the scan sheets. When sheets with bent or torn edges jam in the optical scanner, they are usually mutilated requiring the operator to transcribe the response to a new sheet. A related third point is that the "optical scanner" actually scans in the infrared part of the spectrum. Number two pencils make marks which are detected very well, but even the darkest of marks made with felt tip pens are totally invisible to the scanner. Again, when this type of violation occurs and is detected, the optical

scan operator must remark all responses in the entire booklet using a number 2 pencil. Finally, stray marks on the forms can also cause the machine to halt. The most sensitive areas of the form are usually the top and the timing edges. Oddly enough, job incumbents appear to be rather co-operative in this area, but, for some reason, the survey proctor or administrators do not seem to believe that this prohibition applies to them. Many times such comments as "Batch 1" or "These are all the booklets" have to be erased or snow-painted before the machine will process them.

C. CODAP: PROGRAMMER/ANALYSTS

The programmer/analysts who are responsible for the CODAP system have two observations which may be of interest. These observations may be summarized as the following cautions:

1. Avoid ambiguous coding conventions, and

2. Avoid low response density when possible.

Under certain conditions, a computer program may interpret both blanks and zeroes as zeroes which can be disasterous if separate meanings are intended. This problem has occurred in soliciting a YES/NO response in which a ZERO represented "yes" and a non-response (BLANK) meant "no". This YES/NO choice is usually employed to identify "equipment used" or "courses taken" from a large list. Although using ONE for "yes" and ZERO for "no" would be ideal since it also avoids the low response density, to reduce the marking required for very long lists, it is acceptable to use a non-response (BLANK) for "no". In this way the ambiguity may be used to an advantage. By computing the mean value of ONE and BLANK (zero), the resulting value may be immediately interpreted as "percent of the group" responding yes by simply moving the decimal point two places to the right.

As an example of how both these cautions can be violated in one design step, consider the use of multiple variables to determine a single attribute. This probelm actually occurred in a survey where a 66-valued attribute (coded job title) was solicited on a standard 80 column opscan sheet. Since the scan sheet had 80 columns, each of which had possible responses of BLANK and 0 to 9, the decision was made to solicit the coded job title as seven separate responses. For each of these seven questions, job titles were associated with each of the values 0 through 9, and a non-response (BLANK) indicated "none of the above" for this category. Because of this decision, severe processing problems have been created for the CODAP applications programmers. For each of the seven categories, the BLANK and ZERO responses are indistinguishable for certain programs. Consequently,

the "ZERO" responses for each category includes not only the people who actually responded with a ZERO, but also every person who left that column blank while marking one of the other six. Moreover, because the incumbent was responding to only one out of seven, identically coded columns (0-9), he had the additional problem of locating the proper column for his response. If he inadvertently responded in the wrong column, it is not likely that he caught the mistake because he expected "some" blank columns on either side of his response.

In this case the response density could have been increased by collapsing the seven questions into a single question with a two digit response. By coding the job titles from 01 through 66, the incumbent would have to fill in both columns, maintaining a high response density, and at the same time eliminating the confusion caused by having seven categories to determine a single attribute. From a processing standpoint it is much easier to identify and select subsamples when only one variable needs to be considered.

It is important to note that this category confusion by the incumbent can be caused without actually separating the attribute into distinct questions. In another case, a three column display of supposedly mutually exclusive choices was given in the questionnaire with a corresponding three column response grid on the opscan answer sheet. Some incumbents interpreted this configuration to imply that multiple responses were acceptable and, by stretching their imaginations a little, qualified themselves to fill in the bubbles in two or three columns. Unfortunately, the opscan conversion program was told that the three columns represented only a single response and every multiple response was marked invalid. Although the survey analyst was able to formulate a decision logic for resolving the multiple responses, the optical scan conversion program had already collapsed the multiple responses down to a single asterisk, regardless of which codes had caused the conflict. By the time the tape reached the CODAP applications programmers, the data needed to apply that decision logic had already been lost, and the deadlines precluded any re-scanning efforts.

D. CODAP: APPLICATION PROGRAMMERS

The application programmers who process CODAP studies asked us to reiterate the need for improved communications. The problems mentioned above could have been avoided or minimized had communication been better, but the following example is one in which the lack of communication caused the problem. This classic example comes from a survey in which "length of time on the job" was being solicited in months as a three digit response, but the printed answer sheet only allowed two digits. This error was detected after the forms had been printed but before the inventory was administered. The survey designer

simply instructed the administrators to have the job incumbents record their responses in three previously unused columns at the end of the answer sheet. This would seem to have solved the problem. The decision to change the columns on the answer sheet, however, was not relayed to the person responsible for preparing the data for input to the CODAP system. In the absence of directions to do otherwise, he used the two original columns specified on the form. Later, having noticed the discrepancy, he assumed that the data had effectively been lost due to the fact that people with more than 99 months on the job could not record their response. In talking with the project director one day, he mentioned the loss of the data, and only then was he informed of the additional instructions. By this time several hundred dollars had been spent in processing this data and it was, therefore, prohibitive in cost to simply start over to recover the missing information. The data was eventually recovered after special processing and several days of delay. Luckily, in this case, the opscan system used punched cards for all columns on the answer sheet. There are systems which only record data for the specified areas. Had one of those systems been used in the last example, all the opscan sheets would have had to been reprocessed, either through the optical scanner or manually by keypunch personnel.

As part of this improved communication, the CODAP applications programmers would like to have the opportunity for a post-printing, pre-administration review of intended surveys. In conjunction with the optical scan operators test batches, the applications personnel can help avoid unintentional errors in content. Many times they will encounter one or two duplicated task statements or a task statement which is missing from an inventory which is supposed to exactly match another. These errors can be detected by the computer in the process of automating the titles, and, if this is done prior to the administration of the survey, the incumbents can be alerted to the problem and instructed to take appropriate actions.

IMPLEMENTATION OPTIONS

With all of these points in mind, let us now review several implementation strategies. Having observed several implementations of the optical scan technology, we have identified three general approaches in the occupational analysis area. The first, and least expensive type, employs a standard "off the shelf" 80 column opscan response sheet and a separate questionnaire booklet. The questionnaire booklet indicates the page and column number in which to record both control and response information. The second type also employs a separate questionnaire, but uses a specially designed opscan response booklet in which the number of the response column is the same as the number of the task in the inventory. Background and control information are usually recorded in an area distinct from the normal

task responses. The third approach is to use a single booklet in which the task titles are overprinted directly on the opscan form, adjacent to the appropriate response column. Again, as in type two, as much control information as possible (pages, codes, etc.) is pre-recorded in the over-printing process, but whatever control or background information the incumbent is responsible for, is recorded in an area separated and distinct from the task responses.

A. COSTS

The first approach entails the least direct cost. A standard form is used which is relatively inexpensive and easy to obtain. The printing of the corresponding questionnaire booklet requires no outstanding technical considerations. The second approach also uses the straightforward questionnaire booklet, but is uses specially designed opscan response booklets rather than standard forms. these specially designed opscan forms incur additional expense not only for the initial design costs, but also for the recurring costs of producing forms according to stringent technical requirements. These opscan forms are printed almost exclusively by manufacturers of optical scanning equipment and the major forms producing companies. In the third approach, the separate questionnaire booklet is dropped, but instead the blank opscan forms are sent to a local printer who overprints the control information, instructions, and task statements, then cuts and binds the sheets to form a single document. The local printer who is responsible for printing the task titles and page codes on the opscan forms must be aware of the stringent requirements on these forms. Not only must the critical spacing between page codes be maintained, but after the sheets are printed, care must be taken to avoid cutting or trimming the control edges on these forms. A slight trim on the wrong edge of the sheet can render entire booklets unprocessable. If these sheets are bound together with staples rather than glue, the machines which sever the binding for entry into the optical scanner may experience many nicked or dulled blades. The only safe way to insure that these type three booklets are acceptable is to actually process a sample batch before the surveys are mass produced and sent out into the field.

Aside from these direct costs, there are indirect costs to be considered. Generally speaking, a more sophisticated questionnaire format will require less time and trouble for a job incumbent to complete. Hopefully, this increase in direct cost buys an increase in reliability and a reduction in man-hours required of the organization being surveyed. For example, when using an "off the shelf" 80 column opscan sheet, respondents are directly responsible for coding their booklet number and sheet number on each page. An error in this control information will not be detected by the optical scan processing, but it will cause the entire case to be dropped from any CODAP analysis. Particularly in CODAP studies where job-typing is to be done, a

systematic sampling error can be introduced if a certain group or groups of people tend to make errors in control information. In the Air Force version of CODAP there is a program which is devoted to detecting and identifying such errors so that the data may be corrected whenever possible.

B. OPPORTUNITY FOR TRANSCRIPTION ERRORS

On a type three booklet, once the incumbent selects his response, he merely records it in the adjacent column with no searching required. However, on types one and two, the incumbent, after deciding on his response, must search separate answer sheets for the proper location in which to record his response. For a type two booklet, this search is a simple matter of finding the column number which corresponds directly to the task number. We are assuming, of course, that the users of CODAP have stopped numbering the tasks from 1 to N within each duty, and begun using the new system of numbering the tasks from 1 through the number of tasks in the full inventory, regardless of duty. To respond to a type one booklet, the incumbent is required to locate a particular coding sheet and a particular column whose number may never be the same as the task number.

C. A CASE HISTORY

A "pure" type one approach is rarely used. We have, for example, seen a modified type one approach in which the control information was pre-coded by clerical personnel, and the coded sheets were bound together to form a booklet. In addition, for the few coding forms on which background information was solicited, titles were printed over the columns reserved for those items. With the exception of the fact that the forms still had the standard 1 to 80 numbering on each sheet, this implementation approximately a type two approach. This difference, however, is still significant, particularly when the control information is coded in the first five columns, leaving columns numbered 6 to 77 available for responses.

In an attempt to develop a quantitative measure of the transcription or "parallax" problem, we designed a series of programs to evaluate four inventories. These four, modified type one surveys solicited both relative time spent and a secondary factor. These responses were to be coded on separate sets of answer sheets and each set was to be completed before starting the next. By the definition of a secondary factor, if an incumbent indicated a relative time spent for any task, he was required to respond to the secondary factor for that task also. Because of this, we were provided with a machine sensible way to detect discrepancies. We processed these inventories, listing out all tasks which were missing one of the two responses. In this initial step no attempt was made to distinguish errors of omission from the errors of transcription. The data from this

first step are recorded in Figure 2. This data implies that about one out of every four respondents are making errors and that, for these respondents, the errors account for 5 to 8% of all the information that they are reporting. It is difficult to say whether these figures represent the upper or lower limits of the problem. To the extent that the errors are caused by isolated transcription errors, the figures may be slight over-estimates. Since an error is defined as a task with one rating missing, a person who meant to respond to only two tasks, but misplaced the secondary rating, would show an error rate of 66%, rather than 50% as would be expected. Conversely, to the extent that the errors are caused by blocks of transcription errors, the figures may be under-estimates. In one instance a block of 36 consecutive responses were found in which all the secondary responses were shifted down one column. Is the error rate here actually 5.5% as the algorithm indicates, or is it really 100%? Note also that only discrepancies were detected. Any errors which were systematic went undetected, but systematic errors should be uncommon, correct?

While we were attempting to distinguish between missing responses and shifted responses, several cases were dropped from consideration because of their extremely high error rate and the apparent senselessness of the errors. After all the other cases had been processed, we returned to investigate these cases. Most of them turned out to have a shift error of not one, but six columns. We could not explain why there would be a six column shift among the responses, but we did note that this six column error was almost exclusively restricted to the first opscan coding sheet. At first we surmised that the respondents had simply failed to record the control information in the first five columns and recorded their task responses beginning in column one. This would only account for a shift of five, however. This hypothesis was partially untenable due to the fact that CODAP would drop any case which had an error in the control information, and all the cases studied had already been processed in CODAP. Upon considering the input to CODAP, we noticed that the control information only allowed three digits for case identification and CODAP required four. We then found that some pre-processing had been performed on the data to expand the case control number to four digits. This meant that the control information could account for the entire six-column shift. It was at this point that we learned that the control information had been pre-coded and the incumbent never needed to concern himself with the first five columns of his answer sheet. For someone who did not have to respond to the first five tasks, it seemed only natural to mark his response in the corresponding column number. These "natural instructions" fail when the second coding form is reached which is also numbered 1 to 80. When all else failed, he read and followed the actual instructions. Because of the low response density a respondent expects "some" number of blank columns around his answers and he loses the visual keys or feedback that would help him to detect and correct his errors.

In our analysis of this problem, less than 1% of the people in three of the four surveys had this six-column shift. In the fourth survey, however, this error accounted for over 12% of the population. It should be stated again, these were only the detectable discrepancies. How many people made this error on the first set, realized the mistake while coding the second set and at that point corrected set one? How many people followed their "natural instructions" for set two also, and went undetected? For the people who discovered the error because of the second set, what if there had not been a set two? There are perhaps more questions here than there are answers, but if one values reliability, these are the questions which must be answered.

Perhaps the above problem could be solved by coding the control information in the last five columns of each answer sheet. That way not only could the incumbents use their "natural instructions" for the first 70 or so tasks, if they attempted to respond to task 75, they would find the pre-coded control information there and possibly decide to read the directions.

CLOSING

We hope that we have provided information that may be of use in the design decisions for job survey instruments. We've tried to address those areas in which improved design or communication could have a favorable impact, but there will always be some areas over which we have no control. We, of data processing, have felt that we had no control over the design of the opscan forms, but even lacking the direct control, communication can help solve the problems. We are the ones who experience the consequences of your design decisions, and we feel it is our responsibility to provide feedback in order that your future designs may build upon this knowledge.

JOB SURVEY INFORMATION SYSTEMS

I. SURVEY INSTRUMENT CONSTRUCTION

- A. CONTENT SPECIALIST
- B. DATA COLLECTION INSTRUMENT SPECIALIST
- C. OPTICAL SCAN SHEET MANUFACTURER
- D. LOCAL PRINTER

II. SURVEY ADMINISTRATION

- A. PROCTOR OR ADMINISTRATOR
- B. JOB INCUMBENT

III. SURVEY DATA PROCESSING

- A. PROGRAMMER/ANALYST FOR OPTICAL SCAN MACHINE
- B. OPERATOR OF THE OPTICAL SCAN MACHINE
- C. PROGRAMMER/ANALYST FOR COMPUTER ANALYSIS PROGRAMS
- D. APPLICATIONS PROGRAMMER

IV. SURVEY ANALYSIS

- A. SURVEY ANALYST
- B. RESEARCH ANALYST

Figure 1

ERROR ANALYSIS OF FOUR, MODIFIED TYPE ONE, JOB SURVEYS

	Total Sample - Average Percent of Tasks in error	Only Cases with Errors - Average Percent of Tasks in error	Percent of All cases which contain any errors	Number of Cases with any error	Number of Cases in the Total Sample	Average Number of Tasks Performed	Number of Tasks in the Inventory
SURVEY A	357	36	368	94	25.5	7.81	1.99
SURVEY B	675	56	596	130	21.8	4.99	1.09
SURVEY C	418	62	860	282	32.7	4.83	1.58
SURVEY D	450	91	114	96	84.2	7.87	6.63

Figure 2